

# Stable Cosparse Recovery via $\ell_q$ -analysis Optimization

Shubao Zhang

College of Computer Science & Technology, Zhejiang University  
bravemind@zju.edu.cn

## Abstract

In this paper we study the  $\ell_q$ -analysis optimization ( $0 < q \leq 1$ ) problem for cosparse signal recovery. Our results show that the nonconvex  $\ell_q$ -analysis optimization with  $q < 1$  has better properties than the convex  $\ell_1$ -analysis optimization. In addition, we develop an iteratively reweighted method to solve this problem under the variational framework. Theoretical analysis demonstrates that our method is capable of pursuing a local minima close to the global minima. The empirical results show that the nonconvex approach performs better than its convex counterpart. It is also illustrated that our method outperforms the other state-of-the-art methods for cosparse signal recovery.

## 1 Introduction

The sparse signal recovery problem is widely studied in many areas including compressive sensing [8], statistical estimation [22,23], image processing [13] and signal processing [20]. Traditionally, this problem can be defined from the following under-determined linear equation system

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \quad (1)$$

where  $\mathbf{A} \in \mathbb{R}^{m \times d}$  ( $m \ll d$ ) is the measurement matrix,  $\mathbf{x} \in \mathbb{R}^d$  is the original signal,  $\mathbf{e} \in \mathbb{R}^d$  is a noise vector, and  $\mathbf{y} \in \mathbb{R}^m$  is the noisy observation. Additionally,  $\mathbf{x}$  is assumed to be sparse, or has a sparse representation in some transform domain (that is,  $\mathbf{x} = \mathbf{D}\mathbf{z}$  with a redundant dictionary  $\mathbf{D} \in \mathbb{R}^{d \times n}$  ( $n \gg d$ ) and a sparse vector  $\mathbf{z}$ ). Such an assumption is the basis of the well-known *sparse synthesis model* [7].

The signal  $\mathbf{x}$  can also be supposed to be sparse under a linear transformation. That is, there exists an operator  $\mathbf{\Omega} \in \mathbb{R}^{p \times d}$  ( $p \gg d$ ) such that  $\mathbf{\Omega}\mathbf{x}$  is sparse. This signal model is called the *cosparse analysis model*<sup>1</sup> [16]. Recently, the cosparse analysis model has been demonstrated to be effective and even superior

<sup>1</sup>The synthesis and analysis models are the same when  $\mathbf{D} = \mathbf{\Omega}^{-1}$  provided that both  $\mathbf{D}$  and  $\mathbf{\Omega}$  are square and invertible.

over the synthesis one in many application problems such as signal denoising [9], computer vision [14], signal and image restoration [20], etc.

A natural representation for the cospase signal recovery problem is

$$\min_{\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon. \quad (2)$$

However, this is usually NP-hard due to its combinatorial nature. An alternative is to relax the  $\ell_0$  norm with the  $\ell_1$  norm,

$$\min_{\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon. \quad (3)$$

This model is proposed by Candès *et al.* [4] firstly and called the  $\ell_1$ -analysis minimization. Its equivalent unconstrained formulation is then called the generalized lasso [23]:

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{\Omega}\mathbf{x}\|_1. \quad (4)$$

The  $\ell_1$ -analysis minimization or generalized lasso problem has several special cases, for example, the the total variation (TV) minimization [18], the 2D fused lasso [22], the LLT model [26], and the inf-convolution model [27].

In the statistical estimation literature, the seminal work of Fan and Li [10] showed that the nonconvex sparse recovery problem holds better properties than the convex one. In addition, the nonconvex  $\ell_q$ -synthesis minimization ( $0 < q < 1$ ) problem has been demonstrated to perform better than the convex  $\ell_1$ -synthesis minimization problem [11,19]. Motivated by them, this paper investigates the following  $\ell_q$ -analysis minimization ( $0 < q \leq 1$ ) problem

$$\min_{\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_q^q \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon. \quad (5)$$

We study its theoretical properties. Specifically, we provide an upper bound of the estimate error. It shows that exact recovery can be achieved in both  $q < 1$  and  $q = 1$ . Additionally, we propose an iteratively reweighted method to solve this problem. We further prove that our method is capable to obtain a local minima close to the global minima. The empirical results are consistent with the theoretical analysis. They also show that our iteratively reweighted method outperforms the other state-of-the-art methods such as NESTA [1], split Bregman method [3], and iteratively reweighted  $\ell_1$  method [5] for the  $\ell_1$ -analysis minimization problem and the greedy analysis pursuit (GAP) method [16] for the  $\ell_0$ -analysis minimization problem.

## 1.1 Related Theoretical Work

Candès *et al.* [4] firstly studied the  $\ell_1$ -analysis minimization problem. They provided a  $\ell_2$  norm estimate error bounded by  $C_0\epsilon + C_1s^{-1/2}\|\mathbf{\Omega}\mathbf{x} - (\mathbf{\Omega}\mathbf{x})(s)\|_1$  assuming that  $\mathbf{\Omega}$  is a tight frame ( $\mathbf{\Omega}^*\mathbf{\Omega} = \mathbf{I}$ ). Nam *et al.* [16] studied the  $\ell_1$ -analysis minimization problem without noise. They assumed that every  $p$  rows of  $\mathbf{\Omega}$  are linearly independent. Then they provided a sufficient condition

similar to the exact recovery condition (ERC) which guarantees the exact recovery. Needell and Ward [17] investigated the total variation minimization. They proved that the TV minimization can stably recover a 2D signal. Tibshirani *et al.* [23] proposed the generalized lasso, which includes the lasso as a special case. They studied the property of its solution path. Vaiter *et al.* [24] conducted a robustness analysis of the generalized lasso against noise. Liu *et al.* [15] derived an estimate error bound for the generalized lasso under the assumption that the condition number of  $\mathbf{\Omega}$  is bounded. Specifically, a  $\ell_2$  norm estimate error bounded by  $C\lambda + \|(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \boldsymbol{\epsilon}\|_2$  is given, where  $\boldsymbol{\epsilon}$  is a noise vector and  $C$  is a constant.

## 1.2 Notations

The  $i$ -th entry of a vector  $\mathbf{x}$  is denoted by  $x_i$ . The  $i$ -th row of a matrix  $\mathbf{A}$  is denoted by  $\mathbf{a}_i$ . The best  $k$ -term approximation of a vector  $\mathbf{x} \in \mathbb{R}^d$  is obtained by setting its  $d - k$  insignificant components to zero and denoted by  $\mathbf{x}(k)$ .  $\sigma_{\max}(\mathbf{\Omega})$  and  $\sigma_{\min}(\mathbf{\Omega})$  denote respectively the maximal and minimal singular value of  $\mathbf{\Omega}$ . The  $\ell_q$  norm of a vector  $\mathbf{x} \in \mathbb{R}^d$  is defined as  $\|\mathbf{x}\|_q = (\sum_{i=1}^d |x_i|^q)^{1/q}$ <sup>2</sup> for  $0 < q < \infty$ .  $\lfloor \cdot \rfloor$  denotes the rounding down operator.  $\mathbb{N}$  denotes the natural number.

We now introduce some concepts that characterize the cosparsity analysis model. Instead of the sparsity in the synthesis model that emphasizes the number of few nonzeros of  $\mathbf{z}$ , the analysis model emphasizes on the number of zeros in the representation  $\mathbf{\Omega}\mathbf{x}$  referred to as *cosparsity*. The cosparsity of a vector  $\mathbf{x}$  with respect to  $\mathbf{\Omega}$  is defined as  $l := p - \|\mathbf{\Omega}\mathbf{x}\|_0$ . Such a signal  $\mathbf{x}$  is said to be  $l$ -cosparse. The *support* of a vector  $\mathbf{x}$  is the collection of indices of nonzeros in the vector, denoted by  $T := \{i : x_i \neq 0\}$ . In contrast, the indices of zeros in the representation  $\mathbf{\Omega}\mathbf{x}$  is defined as the *cosupport* of a vector  $\mathbf{x}$ , and denoted by  $\Lambda := \{j : \langle \boldsymbol{\omega}_j, \mathbf{x} \rangle = 0\}$  with  $\boldsymbol{\omega}_j$  the  $j$ -th row of  $\mathbf{\Omega}$ . The submatrix  $\mathbf{\Omega}_\Lambda$  consists of rows in  $\mathbf{\Omega}$  whose indices belong to  $\Lambda$ . Based on these concepts, we can see that a  $l$ -cosparse vector lies in the subspace  $\Sigma_l := \{\mathbf{x} : \mathbf{\Omega}_\Lambda \mathbf{x} = 0, |\Lambda| = l\}$ .

## 2 Exact Recovery via $\ell_q$ -analysis Minimization

In our analysis, we use the  $\mathcal{A}$ -RIP [2].

**Definition 1 ( $\mathcal{A}$ -restricted isometry property)** A matrix  $\Phi$  obeys the  $\mathcal{A}$ -restricted isometry property with constant  $\delta_{\mathcal{A}}$  over any subset  $\mathcal{A} \in \mathbb{R}^N$ , if  $\delta_{\mathcal{A}}$  is the smallest quantity satisfying

$$(1 - \delta_{\mathcal{A}})\|\mathbf{v}\|_2^2 \leq \|\Phi\mathbf{v}\|_2^2 \leq (1 + \delta_{\mathcal{A}})\|\mathbf{v}\|_2^2$$

for all  $\mathbf{v} \in \mathcal{A}$ .

---

<sup>2</sup> $\|\mathbf{x}\|_q$  for  $0 < q < 1$  is not a norm, but  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_q^q$  is a metric.

We can see that the RIP [6], D-RIP [4] and  $\mathbf{\Omega}$ -RIP [12] are all instances of the  $\mathcal{A}$ -RIP with different choices of the set  $\mathcal{A}$ . For example, when choosing  $\mathcal{A} = \{\mathbf{v} : \|\mathbf{v}\|_0 \leq k\}$  and  $\mathcal{A} = \{\mathbf{D}\mathbf{v} : \|\mathbf{v}\|_0 \leq k\}$ , the  $\mathcal{A}$ -restricted isometry are the RIP and D-RIP respectively. It has been verified that any random matrix  $\mathbf{\Phi}$  holds the  $\mathcal{A}$ -restricted isometry property with overwhelming probability provided that the number of measurements depends logarithmically on the number of subspaces in  $\mathcal{A}$  [2].

**Theorem 1** *Assume that the analysis operator  $\mathbf{\Omega}$  has full column rank and the measurement matrix  $\mathbf{A}$  satisfies the  $\mathcal{A}$ -RIP over the set  $\mathcal{A} = \{\mathbf{\Omega}^+\mathbf{v} : \|\mathbf{v}\|_0 \leq k\}$ . Let  $\rho \in \frac{1}{S}\mathbb{N}$  ( $\rho \geq 2$  and  $S \in \mathbb{N}$ ). If the condition number of  $\mathbf{\Omega}$  obeys  $\kappa < \frac{\rho^{1/q-1/2}}{2^{1/q}}$  and the following condition holds*

$$\delta_{\rho S} + (\kappa^{-q}\rho^{1-q/2} - 1)^{2/q}\delta_{(\rho+1)S} < (\kappa^{-q}\rho^{1-q/2} - 1)^{2/q} - 1, \quad (6)$$

*then the minimizer  $\bar{\mathbf{x}}$  of the  $\ell_q$ -analysis minimization problem (5) satisfies*

$$\|\mathbf{x} - \bar{\mathbf{x}}\|_2^q \leq C_1 2^q \varepsilon^q + C_2 \frac{\|\mathbf{\Omega}\mathbf{x} - (\mathbf{\Omega}\mathbf{x})_S\|_q^q}{S^{q/2-1}}$$

where

$$C_1 = \frac{1}{(1 - \delta_{(\rho+1)S})^{q/2}(1 - \kappa^q \rho^{q/2-1}) - \kappa^q \rho^{q/2-1}(1 + \delta_{\rho S})^{q/2}},$$

$$C_2 = \frac{2\sigma_{\min}^{-q}(\mathbf{\Omega})\rho^{q/2-1}}{1 - \kappa^q \rho^{q/2-1}}[C_1(1 + \delta_{\rho S})^{q/2} + 1].$$

This theorem says that the  $\ell_q$ -analysis optimization can stably recover the cospase signals under certain conditions. Especially, exact recovery can be achieved in the noiseless case for all signals  $\mathbf{x}$  such that  $\mathbf{\Omega}\mathbf{x}$  is  $S$ -sparse. It is easy to check that the constant  $C_1$  is monotonically increasing with respect to  $q \in (0, 1]$ . This implies that the case  $q < 1$  is more robust to noise than the case  $q = 1$  because of a tighter error bound.

A slightly stronger condition than the condition (6) is

$$\delta_{(\rho+1)S} < \frac{(\kappa^{-q}\rho^{1-q/2} - 1)^{2/q} - 1}{(\kappa^{-q}\rho^{1-q/2} - 1)^{2/q} + 1}. \quad (7)$$

It is easy to verify that the function  $(\kappa^{-q}\rho^{1-q/2} - 1)^{2/q}$  is monotonically decreasing with respect to  $q \in (0, 1]$ . Therefore, the above condition is less restrictive for a smaller  $q$ . Given a  $\rho$ , a larger condition number  $\kappa$  will makes the above condition more restrictive, because the value of the inequality's right-hand side becomes smaller. In other words, an analysis dictionary with a too large condition number could let the  $\ell_q$ -analysis minimization fail to do recovery. This provides some hints on the selection of the analysis dictionary. The right-hand side of the above inequality tends to infinity as  $q$  approaches zero. The following result is then straightforward.

**Proposition 1** *Assuming that  $\delta_{(\rho+1)S} < 1$ , then all  $(p - S)$ -cosparse signals can be exactly recovered via the  $\ell_q$ -analysis minimization for some small enough  $q > 0$ .*

The condition (7) guarantees that all  $(p - S)$ -cosparse signals can be recovered via the  $\ell_q$ -analysis minimization. Define  $S_q$  ( $0 < q \leq 1$ ) as the largest value of the sparsity  $S \in \mathbb{N}$  of the analysis coefficient  $\mathbf{\Omega}\mathbf{x}$  such that the condition (7) holds for some  $\rho \in \frac{1}{S}\mathbb{N}$  ( $\rho \geq 2$ ). The following proposition indicates the relationship between  $S_q$  with  $q < 1$  and  $S_1$  with  $q = 1$ .

**Proposition 2** *Suppose that there exist  $S_1 \in \mathbb{N}$  and  $\rho_1 \in \frac{1}{S_1}\mathbb{N}$  ( $\rho_1 \geq 2$ ) such that*

$$\delta_{(\rho_1+1)S_1} < \frac{(\kappa^{-1}\rho_1^{1/2} - 1)^2 - 1}{(\kappa^{-1}\rho_1^{1/2} - 1)^2 + 1}.$$

*Then there exist  $\rho_q \in \frac{1}{S_q}\mathbb{N}$  ( $\rho_q \geq 2$ ) and  $S_q \in \mathbb{N}$  obeying*

$$S_q = \left\lfloor \frac{\rho_1 + 1}{\rho_1^{\frac{1}{2-q}} + 1} \right\rfloor S_1 \quad (8)$$

*such that  $(\rho_1 + 1)S_1 = (\rho_q + 1)S_q$  and*

$$\delta_{(\rho_q+1)S_q} < \frac{(\kappa^{-q}\rho_q^{1-q/2} - 1)^{2/q} - 1}{(\kappa^{-q}\rho_q^{1-q/2} - 1)^{2/q} + 1}.$$

The equation (8) states that the  $\ell_q$ -analysis minimization with  $q < 1$  can do recovery in a wider range of cosparsity than the  $\ell_1$ -analysis minimization.

### 3 Iteratively Reweighted Method for Cosparse Recovery

We consider the following unconstrained optimization problem

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{\Omega}\mathbf{x}\|_q^q \right\}. \quad (9)$$

It is difficult to directly tackle the above problem because of the nonsmoothness and nonseparability of the  $\ell_q$  norm term. We provide a way to deal with the  $\ell_q$  norm under the variational framework.

Note that the function  $\|\mathbf{\Omega}\mathbf{x}\|_q^q$  is concave with respect to  $|\mathbf{\Omega}\mathbf{x}|^\alpha = (|\omega_{1,x}|^\alpha, \dots, |\omega_{d,x}|^\alpha)^T$  for  $\alpha \geq 1$ . Thus there exists a variational upper bound of  $\|\mathbf{\Omega}\mathbf{x}\|_q^q$ . Given a positive vector  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_d)^T$ , we consider the following minimization problem

$$\|\mathbf{\Omega}\mathbf{x}\|_q^q = \sum_{i=1}^d (|\omega_{i,x}|^\alpha)^{\frac{q}{\alpha}} = \min_{\boldsymbol{\eta} > \mathbf{0}} \left\{ J_\alpha = \frac{q}{\alpha} \sum_{i=1}^d \left( \eta_i |\omega_{i,x}|^\alpha + \frac{\alpha - q}{q} \frac{1}{\eta_i^{\frac{q}{\alpha-q}}} \right) \right\}$$

for  $\alpha \geq 1$  and  $0 < q \leq 1$ . The function  $J_\alpha$  is jointly convex in  $(\mathbf{x}, \boldsymbol{\eta})$ . Its minimum is achieved at  $\eta_i = 1/|\boldsymbol{\omega}_i \mathbf{x}|^{\alpha-q}$ ,  $i = 1, \dots, d$ . However, care must be taken when  $\mathbf{x}$  involves some zero coordinates. In such a case, the weight vector  $\boldsymbol{\eta}$  would include infinite components. To avoid infinite weight, we add a smoothing term  $q/\alpha \sum_{i=1}^d \eta_i \varepsilon^\alpha$  ( $\varepsilon \geq 0$ ) to  $J_\alpha$ .

Using the above variational formulation, we can reformulate the unconstrained problem (9) as

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}, \varepsilon) = \min_{\boldsymbol{\eta} > \mathbf{0}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\lambda q}{\alpha} \sum_{i=1}^p \left[ \eta_i (|\boldsymbol{\omega}_i \mathbf{x}|^\alpha + \varepsilon^\alpha) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{\frac{q}{\alpha-q}}} \right] \right\}. \quad (10)$$

We then develop an alternating minimization algorithm consisting of three steps. The first step calculates  $\boldsymbol{\eta}$  with  $\mathbf{x}$  fixed via

$$\boldsymbol{\eta}^{(k)} = \operatorname{argmin}_{\boldsymbol{\eta} > \mathbf{0}} \left\{ \sum_{i=1}^p \left[ \eta_i (|\boldsymbol{\omega}_i \mathbf{x}|^\alpha + \varepsilon^\alpha) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{\frac{q}{\alpha-q}}} \right] \right\},$$

which has a closed form solution. The second step calculates  $\mathbf{x}$  with  $\boldsymbol{\eta}$  fixed via

$$\mathbf{x}^{(k)} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^d} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\lambda q}{\alpha} \sum_{i=1}^p \eta_i |\boldsymbol{\omega}_i \mathbf{x}|^\alpha \right\},$$

which is a weighted  $\ell_\alpha$ -minimization problem. Particularly, the case  $\alpha = 2$  corresponds to a least squares problem which can be solved efficiently. The third step updates the smoothing parameter  $\varepsilon$  according to the following rule<sup>3</sup>

$$\varepsilon^{(k)} = \min\{\varepsilon^{(k-1)}, \rho \cdot r(\boldsymbol{\Omega} \mathbf{x}^{(k)})_l\} \quad \text{with } \rho \in (0, 1),$$

where  $r(\boldsymbol{\Omega} \mathbf{x})_l$  is the  $l$ -th smallest element of the set  $\{|\boldsymbol{\omega}_j \mathbf{x}| : j = 1, \dots, p\}$ .  $\mathbf{x}$  is a  $l$ -cosparsely vector if and only if  $r(\boldsymbol{\Omega} \mathbf{x})_l = 0$ . The algorithm stops when  $\varepsilon = 0$ .

### 3.1 Convergence Analysis

Our analysis is based on the optimization problem (10) with the objective function  $F(\mathbf{x}, \varepsilon)$ . Noting that  $\boldsymbol{\eta}^{(k+1)}$  is a function of  $\mathbf{x}^{(k)}$  and  $\varepsilon^{(k)}$ , we define the following objective function

$$Q(\mathbf{x}, \varepsilon | \mathbf{x}^{(k)}, \varepsilon^{(k)}) \triangleq \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\lambda q}{\alpha} \sum_{i=1}^p \left[ \eta_i^{(k+1)} (|\boldsymbol{\omega}_i \mathbf{x}|^\alpha + \varepsilon^\alpha) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{(k+1) \frac{q}{\alpha-q}}} \right].$$

**Lemma 1** *Assume that the analysis dictionary  $\boldsymbol{\Omega}$  has full column rank. Let  $\{(\mathbf{x}^{(k)}, \varepsilon^{(k)}) : k = 1, 2, \dots\}$  be a sequence generated by the CoIRLq algorithm. Then,*

$$\|\mathbf{x}^{(k)}\|_2 \leq \|(\boldsymbol{\Omega}^T \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^T\|_2 (F(\mathbf{x}^{(0)}, \varepsilon^{(0)})/\lambda)^{1/q}$$

<sup>3</sup>Various strategies can be applied to update  $\varepsilon$ . For example, we can keep  $\varepsilon$  as a small fixed value. It is preferred to choose a sequence of  $\{\varepsilon^{(k)}\}$  tending to zero [25].

---

**Algorithm 1** The CoIRLq Algorithm

---

**Input:**  $l, \mathbf{A}, \mathbf{y}, \mathbf{\Omega} = [\boldsymbol{\omega}_1^T, \dots, \boldsymbol{\omega}_p^T]^T$ .

**Init:** Choose  $\mathbf{x}^{(0)}$  such that  $\mathbf{A}\mathbf{x}^{(0)} = \mathbf{y}$  and  $\varepsilon^{(0)} = 1$ .

**while**  $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_\infty > \tau$  or  $\varepsilon^{(k)} \neq 0$  **do**  
    Update

$$\boldsymbol{\eta}^{(k)} = \underset{\boldsymbol{\eta} > \mathbf{0}}{\operatorname{argmin}} \left\{ \sum_{i=1}^p \left( \eta_i (|\boldsymbol{\omega}_i \mathbf{x}^{(k-1)}|^\alpha + \varepsilon^\alpha) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{\frac{q}{\alpha - q}}} \right) \right\},$$

and the update rule is the following

$$\eta_i^{(k)} = \left( |\boldsymbol{\omega}_i \mathbf{x}^{(k-1)}|^\alpha + \varepsilon^{(k-1)\alpha} \right)^{q/\alpha - 1}.$$

Update

$$\mathbf{x}^{(k)} = \underset{\mathbf{x} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\lambda q}{\alpha} \sum_{i=1}^p \eta_i^{(k)} |\boldsymbol{\omega}_i \mathbf{x}|^\alpha \right\}.$$

Update

$$\varepsilon^{(k)} = \min\{\varepsilon^{(k-1)}, \rho \cdot r(\mathbf{\Omega}\mathbf{x}^{(k)})_l\} \text{ with } \rho \in (0, 1).$$

**end while**

**Output:**  $\mathbf{x}$

---

and

$$F(\mathbf{x}^{(k+1)}, \varepsilon^{(k+1)}) \leq F(\mathbf{x}^{(k)}, \varepsilon^{(k)})$$

with equality if and only if  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$  and  $\varepsilon^{(k+1)} = \varepsilon^{(k)}$ .

The boundedness of  $\|\mathbf{x}^{(k)}\|_2$  implies that the sequence  $\{\mathbf{x}^{(k)}\}$  converges to some accumulation point.

Let  $\mathcal{H}(\mathbf{x}^{(k)}, \varepsilon^{(k)})$  be the set of values of  $(\mathbf{x}, \varepsilon)$  that minimize  $Q(\mathbf{x}, \varepsilon | \mathbf{x}^{(k)}, \varepsilon^{(k)})$  over  $\Omega \subset \mathbb{R}^d \times \mathbb{R}_+$  and  $\mathcal{S}$  be the set of stationary points of  $Q$  in the interior of  $\Omega$ . We can immediately derive the following theorem from Zangwill's *global convergence theorem* or the literature [21].

**Theorem 2** Let  $\{\mathbf{x}^{(k)}, \varepsilon^{(k)}\}$  be a sequence of the CoIRLq algorithm generated by  $(\mathbf{x}^{(k+1)}, \varepsilon^{(k+1)}) \in \mathcal{H}(\mathbf{x}^{(k)}, \varepsilon^{(k)})$ . Suppose that (i)  $\mathcal{H}(\mathbf{x}^{(k)}, \varepsilon^{(k)})$  is closed over the complement of  $\mathcal{S}$  and (ii)

$$F(\mathbf{x}^{(k+1)}, \varepsilon^{(k+1)}) < F(\mathbf{x}^{(k)}, \varepsilon^{(k)}) \quad \text{for all } (\mathbf{x}^{(k)}, \varepsilon^{(k)}) \notin \mathcal{S}.$$

Then all the limit points of  $\{\mathbf{x}^{(k)}, \varepsilon^{(k)}\}$  are stationary points of  $F(\mathbf{x}, \varepsilon)$  and  $F(\mathbf{x}^{(k)}, \varepsilon^{(k)})$  converges monotonically to  $F(\mathbf{x}^*, \varepsilon^*)$  for some stationary point  $(\mathbf{x}^*, \varepsilon^*)$ .

### 3.2 Recovery Guarantee Analysis

To uniquely recover the original signal, the linear operator  $\mathbf{A} : \mathcal{A} \rightarrow \mathbb{R}^m$  must be a one-to-one map. Define a set  $\bar{\mathcal{A}} = \{\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 : \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}\}$ . Blumensath and Davies [2] showed that a necessary condition for the existence of a one-to-one map requires that  $\delta_{\bar{\mathcal{A}}} < 1$  ( $\delta_{\mathcal{A}} \leq \delta_{\bar{\mathcal{A}}}$ ). For any two  $l$ -cosparse vectors  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A} = \{\mathbf{x} : \Omega_{\Lambda} \mathbf{x} = \mathbf{0}, |\Lambda| \geq l\}$ , denote  $T_1 = \text{supp}(\Omega \mathbf{x}_1)$ ,  $T_2 = \text{supp}(\Omega \mathbf{x}_2)$ ,  $\Lambda_1 = \text{cosupp}(\Omega \mathbf{x}_1)$  and  $\Lambda_2 = \text{cosupp}(\Omega \mathbf{x}_2)$ . Since  $\text{supp}(\Omega(\mathbf{x}_1 + \mathbf{x}_2)) \subseteq T_1 \cup T_2$ , we have  $\text{cosupp}(\Omega(\mathbf{x}_1 + \mathbf{x}_2)) \supseteq (T_1 \cup T_2)^C = T_1^C \cap T_2^C = \Lambda_1 \cap \Lambda_2$ . Moreover, we also have  $|\Lambda_1 \cap \Lambda_2| = p - |T_1 \cup T_2| \geq p - (p - l) - (p - l) = 2l - p$ . Thus for the cosparse signal model, it requires that the measurement matrix  $\mathbf{A}$  satisfies the  $\mathcal{A}$ -RIP with  $\delta_{2l-p} < 1$  to uniquely recover any  $l$ -cosparse vector from the set  $\mathcal{A} = \{\mathbf{x} : \Omega_{\Lambda} \mathbf{x} = \mathbf{0}, |\Lambda| \geq l\}$ . Otherwise, there would exist two  $l$ -cosparse vectors  $\mathbf{x}_1 \neq \mathbf{x}_2$  such that  $\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$ . Giryes *et al.* [12] showed that there does exist random matrix  $\mathbf{A}$  satisfying such requirement with high probability.

**Theorem 3** *Suppose that  $\mathbf{x}^*$  is the true signal vector satisfying  $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$  with  $\|\mathbf{e}\|_2 \leq \varepsilon$ . Assume that the measurement matrix  $\mathbf{A}$  satisfies the  $\mathcal{A}$ -RIP property of order  $2l - p$  with  $\delta_{2l-p} < 1$  over the set  $\mathcal{A} = \{\mathbf{x} : \Omega_{\Lambda} \mathbf{x} = \mathbf{0}, |\Lambda| \geq l\}$ . Then the solution  $\bar{\mathbf{x}}$  obtained by the CoIRLq algorithm satisfies the following error bound*

$$\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2 \leq C_1 \sqrt{\lambda} + C_2 \varepsilon,$$

where  $C_1$  and  $C_2$  are constants depending on  $\delta_{2l-p}$ .

We can see that the CoIRLq algorithm can recover an approximate solution away from the true signal vector by a factor of  $\sqrt{\lambda}$  in the noiseless case.

## 4 Numerical Analysis

In this section we conduct a numerical analysis of the  $\ell_q$ -analysis minimization method on both simulated data and real data. We compare the performance of the case  $q < 1$  and the case  $q = 1$ . We set  $\alpha = 2$  in the CoIRLq algorithm. All the codes we used will be published.

### 4.1 Simulated Experiment

We generate the simulated datasets according to the following model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \sigma\boldsymbol{\epsilon},$$

where  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I})$  and  $\sigma$  is the noise level. The measurement matrix  $\mathbf{A}$  is drawn independently from the normal distribution with normalized columns. The analysis dictionary  $\Omega$  is constructed such that  $\Omega^T$  is a random tight frame. To generate a  $l$ -cosparse signal  $\mathbf{x}$ , we firstly choose  $l$  rows randomly from  $\Omega$  and form  $\Omega_{\Lambda}$ . Then we generate a signal which lies in the null space of  $\Omega_{\Lambda}$ .



The recovery is deemed to be successful if the recovery relative error  $\|\bar{\mathbf{x}} - \mathbf{x}^*\|_2 / \|\mathbf{x}^*\|_2 \leq 1e-4$ .

In the first experiment, we test the signal recovery capability of the CoIRLq method with  $q = 0.7$ . We set  $m = 80, p = 144, d = 120, l = 99$ , and  $\sigma = 0$ . Figure 1 illustrates that the CoIRLq method recovers the original signal perfectly.

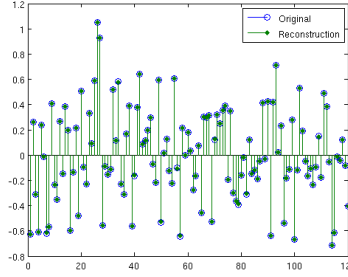
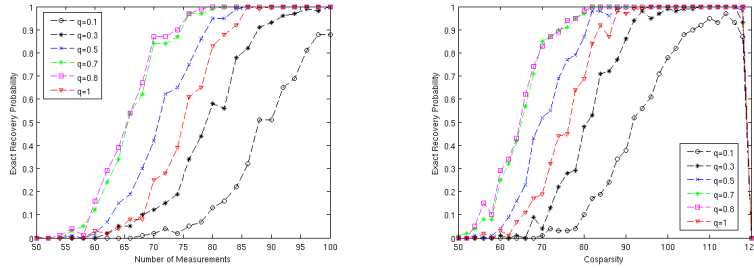


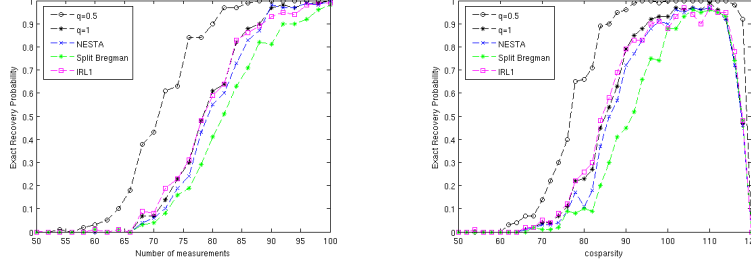
Figure 1: Cosparse signal recovery.

In the second experiment, we test the CoIRLq method on a range of measurement number and cosparsity respectively with different  $q$ . Although the optimal parameter  $\lambda$  generally depends on  $q$ , a small enough  $\lambda$  is able to ensure that  $\mathbf{y}$  approximately equals to  $\mathbf{A}\mathbf{x}$  in the noiseless case. Thus, we set the parameter  $\lambda = 1e-4$ . Figure 2 reports the result with 100 repetitions on every dataset. We can see that the CoIRLq method with  $q = 0.5, 0.7, 0.8$  can achieve exact recovery in a wider range of cosparsity and with fewer measurements than with  $q = 1$ . In addition, it should be noted that a small  $q = 0.1, 0.3$  doesn't perform better than a relatively large  $q = 0.7, 0.8$ , since a too small  $q$  leads to a hard-solving problem.



$$p = 144, d = 120, l = 99, \sigma = 0 \quad m = 90, p = 144, d = 120, \sigma = 0$$

Figure 2: Exact recovery probability of the CoIRLq method.



$$p = 144, d = 120, l = 99, \sigma = 0.01 \quad m = 90, p = 144, d = 120, \sigma = 0.01$$

Figure 3: Exact recovery probability of the CoIRLq, split Bregman, NESTA and IRL1 methods.

In the third experiment, we compare the CoIRLq method with three state-of-the-art methods for the  $\ell_1$ -analysis minimization problem including NESTA<sup>4</sup> [1], Split Bregman method [3], and iteratively reweighted  $\ell_1$  (IRL1) method [5]. The parameter  $\lambda$  is tuned via the grid search method. We run these methods in a range of measurement number and cosparsity. Figure 3 reports the result with 100 repetitions on every dataset. We can see that the nonconvex  $\ell_q$ -analysis minimization with  $q < 1$  is more capable of achieving exact recovery against noise than the convex  $\ell_1$ -analysis minimization. Moreover, the nonconvex approach can obtain exact recovery with fewer measurements or in a wider range of cosparsity than the convex counterpart. Moreover, we found that the CoIRLq algorithm in the case  $q < 1$  often needs less iterations than in the case  $q = 1$ .

## 4.2 Image Restoration Experiment

In this section we demonstrate the effectiveness of the CoIRLq method on the Shepp Logan phantom reconstruction problem. In computed tomography, an image can not be observed directly. Instead, we can only obtain its 2D Fourier transform coefficients along a few radial lines due to certain limitations. This sampling process can be modeled as a measurement matrix  $\mathbf{A}$ . The goal is to reconstruct the image from the observation.

The experimental program is set as follows. The image dimension is of  $256 \times 256$ , namely  $d = 65536$ . The sampling operator  $\mathbf{A}$  is a two dimensional Fourier transform which measures the image's Fourier transform along a few radial lines. The analysis dictionary is a finite difference operator  $\mathbf{\Omega}_{2D-DIF}$  whose size is roughly twice the image size, namely  $p = 130560$ . Since the number of nonzero analysis coefficients is  $p - l = 2546$ , the cosparsity used is  $l = p - 2546 = 128014$ . The number of measurements depends on the number of radial lines used. To show the reconstruction capability of the CoIRLq method,

<sup>4</sup><http://statweb.stanford.edu/~candes/nesta/>

we conduct the following experiments (the parameter  $\lambda$  is tuned via grid search). First, we compare our method with the greedy analysis pursuit (GAP<sup>5</sup>) method for the  $\ell_0$ -analysis minimization (4) [16]. Figure (f), (g) and (h) show that our method performs better than the GAP method in the noisy case. We can see that the CoIRLq method with  $q < 1$  is more robust to noise than with  $q = 1$ . Second, we take an experiment using 10 radial lines without noise. The corresponding number of measurements is  $m = 2282$ , which is approximately 3.48% of the image size. Figure (c) demonstrates that the CoIRLq ( $q = 0.7$ ) method with 10 lines obtains perfect reconstruction. Figure (d) shows that the CoIRLq ( $q = 1$ ) method with 12 lines attains perfect reconstruction. The GAP method however needs at least 12 radial lines to achieve exact recovery; see [16].

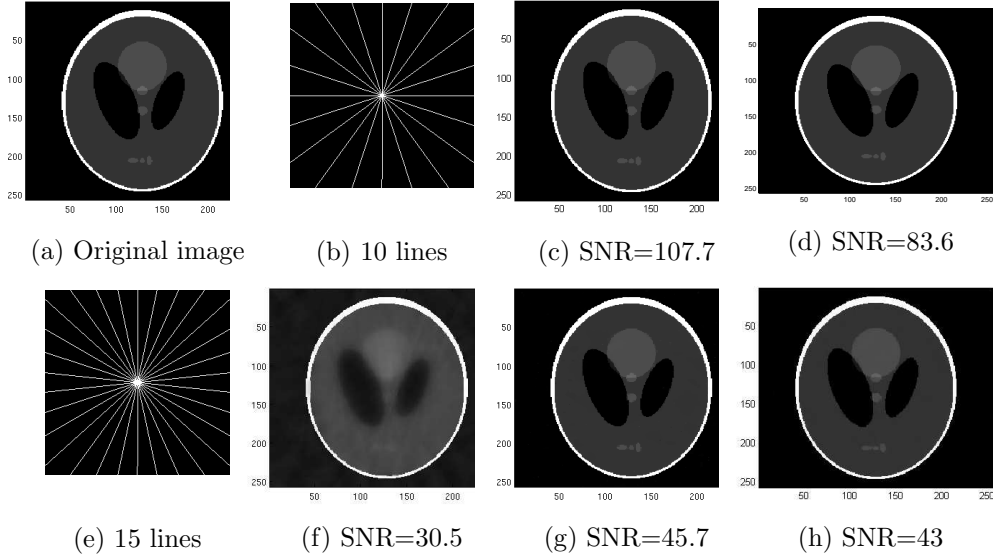


Figure 4: (a) Original Shepp Logan Phantom image; (b) Sampling locations along 10 radial lines; (c) Exact reconstruction via CoIRLq ( $q=0.7$ ) with 10 lines without noise; (d) Exact reconstruction via CoIRLq ( $q=1$ ) with 12 lines without noise; (e) Sampling locations along 15 radial lines; (f) Reconstruction via GAP with 15 lines and noise level  $\sigma = 0.01$ ; (g) Reconstruction via CoIRLq ( $q=0.7$ ) with 15 lines and noise level  $\sigma = 0.01$ ; (h) Reconstruction via CoIRLq ( $q=1$ ) with 15 lines and noise level  $\sigma = 0.01$ .

## References

- [1] Becker, S., Bobin, J., Candès, E.J. (2011) NESTA: A fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences* 4(1), 1–39.

<sup>5</sup><http://www.small-project.eu/software-data>

- [2] Blumensath, T., Davies, M.E. (2008) Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Transactions on Information Theory* **55**(4), 1872–1882.
- [3] Cai, J.F., Osher, S., Shen, Z.W. (2009) Split bregman methods and frame based image restoration. *Multiscale Modeling and Simulation* **8**(2), 337–369.
- [4] Candès, E.J., Eldar, Y.C., Needle, D., Randall, P. (2010) Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis* **31**(1), 59–73.
- [5] Candès, E.J., Wakin, M., Boyd, S. (2007) Enhancing sparsity by reweighted  $\ell_1$  minimization. *J. Fourier Anal. Appl.* **14**, 877–905.
- [6] Candès, E.J., Tao, T. (2004) Decoding by linear programming. *IEEE Transactions on Information Theory* **51**, 4203–4215.
- [7] Donoho, D.L., Elad, M. (2003) Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization. In *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2334–2345.
- [8] Donoho, D.L. (2006) Compressed sensing. *IEEE Transactions on Information Theory* **52**, 1289–1306.
- [9] Elad, M.; Milanfar, P., Rubinstein, R. (2007) Analysis versus synthesis in signal priors. *Inverse Problems* **23**(3), 947–968.
- [10] Fan, J., Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association*.
- [11] Foucart, S., Lai, M.J. (2009) Sparsest solutions of underdetermined linear systems via  $\ell_q$ -minimization for  $0 < q \leq 1$ . *Applied and Computational Harmonic Analysis* **26**, 395–407.
- [12] Giryes, R., Nam, S., Elad, M., Gribonval, R., Davies, M.E. (2013) Greedy-like algorithms for the cospase analysis model. *Special Issue in LAA on Sparse Approximation Solutions of Linear Systems*.
- [13] Krishnan, D., Fergus, R. (2009) Fast image deconvolution using hyper-laplacian priors. In *Advances in Neural Information Processing Systems*.
- [14] Kiechle, M., Hawe, S., Kleinstaub, M. (2013) A joint intensity and depth cospase analysis model for depth map super-resolution. In *CVPR*.
- [15] Liu, J., Yuan, L., Ye, J.P. (2013) Guaranteed sparse recovery under linear transformation. In *Proceedings of the 30th International Conference on Machine Learning*.
- [16] Nam, S., Davies, M.E., Elad, M., Gribonval, R. (2011) The cospase analysis model and algorithms. *Applied and Computational Harmonic Analysis* **34**(1), 30–56.
- [17] Needell, D., Ward, R. (2013) Stable image reconstruction using total variation minimization. *SIAM Journal on Imaging Sciences*, vol.6, no. 2, 1035–1058.
- [18] Rudin, L.I., Osher, S., Fatemi, E. (1992) Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268.
- [19] Saab, R., Yilmaz, Ö. (2010) Sparse recovery by nonconvex optimization - instance optimality. *Applied and Computational Harmonic Analysis* **29**, 30–48.
- [20] Selesnick, I.W., Figueiredo, M.A.T. (2009) Signal restoration with overcomplete wavelet transform: Comparison of analysis and synthesis priors. In *Proceedings of SPIE*.
- [21] Sriperumbudur, B.K., Lanckriet, G.R.G. (2009) On the convergence of the concave-convex procedure. In *Advances in Neural Information Processing Systems* **22**.
- [22] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K. (2005) Sparsity and smoothness via the fused lasso. *J. R. Statist. Soc. B* **67**, Part 1, 91–108.

- [23] Tibshirani, R.J., Taylor, J. (2011) The solution path of the generalized lasso. *The Annals of Statistics*, Vol. 39, No. 3, 1335-1371.
- [24] Vaiter, S., Peyre, G., Dossal, C., Fadili, J. (2013) Robust sparse analysis regularization. *IEEE Transactions on Information Theory* **59**(4), 2001-2016.
- [25] Daubechies, I., Devore, R., Fornasier, M., Güntürk, C.S. (2010) Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*.
- [26] Lysaker, M., Lundervold, A., Tai X.C. (2003) Noise removal using fourth-order partial differential equation with application to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing* **41**(12), 3397-3415.
- [27] Chambolle, A., Lions, P. (1997) Image recovery via total variation minimization and related problems. *Numer. Math.* **76**(2),167-188.

## A: Proof of Theorem 1

**Notations** Let  $\mathbf{h} = \mathbf{x} - \bar{\mathbf{x}}$  ( $\mathbf{x}$  denotes the original signal,  $\bar{\mathbf{x}}$  denotes the estimated signal). Denote the index set of  $\mathbf{\Omega}\mathbf{x}$  by  $T$ .  $T_0$  is the collection of indices corresponding to the largest  $S$  elements of  $\mathbf{\Omega}\mathbf{x}$ . We partition the complementary set  $T_0^c = T/T_0$  into the following subsets without intersection:  $\{T_1, T_2, \dots\}$  such that  $T_1$  indicates the index set of the largest  $M$  elements in  $\mathbf{\Omega}_{T_0^c}\mathbf{h}$ ,  $T_2$  indicates the index set of the next largest  $M$  elements in  $\mathbf{\Omega}_{T_0^c}\mathbf{h}$ , and so forth (the size of the last one may be less than  $M$ ). Particularly, denote  $T_{01} = T_0 \cup T_1$ .  $\mathbf{\Omega}_{T_j}$  is a submatrix of  $\mathbf{\Omega}$  restricted to  $T_j$ .  $|\mathbf{v}|_{(k)}$  denotes the  $k$ -th element of  $|\mathbf{v}| = [|v_1|, |v_2|, \dots]^T$ .  $\sigma_{max}(\mathbf{\Omega})$  denotes the maximal singular value of  $\mathbf{\Omega}$ , while  $\sigma_{min}(\mathbf{\Omega})$  denotes its minimal singular value.  $\lfloor \cdot \rfloor$  denotes the rounding down operator.

**Lemma 2** *The following inequality holds*

$$\|\mathbf{\Omega}_{T_0^c}\mathbf{h}\|_q^q \leq \|\mathbf{\Omega}_{T_0}\mathbf{h}\|_q^q + 2\|\mathbf{\Omega}_{T_0^c}\mathbf{x}\|_q^q.$$

**Proof** Both  $\mathbf{x}$  and  $\bar{\mathbf{x}}$  are the feasible solutions. But  $\bar{\mathbf{x}}$  is the minimizer of the problem (1). Thus we have

$$\begin{aligned} \|\mathbf{\Omega}\mathbf{x}\|_q^q &\geq \|\mathbf{\Omega}\bar{\mathbf{x}}\|_q^q = \|\mathbf{\Omega}\mathbf{x} - \mathbf{\Omega}\mathbf{h}\|_q^q = \|\mathbf{\Omega}_{T_0}\mathbf{x} - \mathbf{\Omega}_{T_0}\mathbf{h}\|_q^q + \|\mathbf{\Omega}_{T_0^c}\mathbf{x} - \mathbf{\Omega}_{T_0^c}\mathbf{h}\|_q^q \\ &\geq \|\mathbf{\Omega}_{T_0}\mathbf{x}\|_q^q - \|\mathbf{\Omega}_{T_0}\mathbf{h}\|_q^q + \|\mathbf{\Omega}_{T_0^c}\mathbf{h}\|_q^q - \|\mathbf{\Omega}_{T_0^c}\mathbf{x}\|_q^q \end{aligned}$$

The second inequality is because of that the pseudonorm  $\|\cdot\|_q^q$  satisfies the triangle inequality. ■

**Lemma 3** *Let  $\eta = \|\mathbf{\Omega}_{T_0^c}\mathbf{x}\|_q^q$ . The following inequalities hold*

$$\|\mathbf{\Omega}_{T_0}\mathbf{h}\|_q \leq S^{1/q-1/2} \|\mathbf{\Omega}_{T_0}\mathbf{h}\|_2$$

and

$$\sum_{j \geq 2} \|\mathbf{\Omega}_{T_j}\mathbf{h}\|_2^q \leq M^{q/2-1} (\|\mathbf{\Omega}_{T_0}\mathbf{h}\|_q^q + 2\eta).$$

**Proof** The first inequality is because of the Hölder's inequality. Now we prove the second inequality. Since any element in  $T_{j+1}$  is less than the average in  $T_j$ , we have

$$|\mathbf{\Omega}_{T_{j+1}}\mathbf{h}|_{(k)}^q \leq \|\mathbf{\Omega}_{T_j}\mathbf{h}\|_q^q / M.$$

Thus we have

$$\|\mathbf{\Omega}_{T_{j+1}}\mathbf{h}\|_2 \leq M^{1/2-1/q} \|\mathbf{\Omega}_{T_j}\mathbf{h}\|_q.$$

Further, we obtain

$$\sum_{j \geq 2} \|\mathbf{\Omega}_{T_j}\mathbf{h}\|_2^q \leq M^{q/2-1} \sum_{j \geq 1} \|\mathbf{\Omega}_{T_j}\mathbf{h}\|_q^q = M^{q/2-1} \|\mathbf{\Omega}_{T_0^c}\mathbf{h}\|_q^q.$$

Using the lemma 2, we get the the second inequality. ■

**Lemma 4** *The following inequality holds*

$$\|\mathbf{A}\mathbf{h}\|_2 \leq 2\varepsilon.$$

**Proof** Since both  $\mathbf{x}$  and  $\bar{\mathbf{x}}$  are feasible, we have

$$\|\mathbf{A}\mathbf{h}\|_2 = \|\mathbf{y} - \mathbf{A}\mathbf{x} - (\mathbf{y} - \mathbf{A}\bar{\mathbf{x}})\|_2 \leq \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \|\mathbf{y} - \mathbf{A}\bar{\mathbf{x}}\|_2 \leq 2\varepsilon.$$

■

**Lemma 5** Assume that the analysis dictionary  $\mathbf{\Omega}$  has full column rank and the matrix  $\mathbf{A}$  satisfies the D-RIP. Then the following holds

$$(1 - \delta_{S+M})^{q/2} \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q \leq (2\varepsilon)^q + \sigma_{\min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} M^{q/2-1} (S^{1-q/2} \|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q + 2\eta).$$

**Proof** Since  $\mathbf{\Omega}$  has full column rank,  $\mathbf{\Omega}^+ \mathbf{\Omega} = \mathbf{I}$  holds. Then we have

$$\begin{aligned} (2\varepsilon)^q &\geq \|\mathbf{A}\mathbf{h}\|_2^q = \|\mathbf{A}\mathbf{\Omega}^+ \mathbf{\Omega} \mathbf{h}\|_2^q \\ &\geq \|\mathbf{A}\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q - \sum_{j \geq 2} \|\mathbf{A}\mathbf{\Omega}_{T_j}^+ \mathbf{\Omega}_{T_j} \mathbf{h}\|_2^q \\ &\geq (1 - \delta_{S+M})^{q/2} \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q - (1 + \delta_M)^{q/2} \sum_{j \geq 2} \|\mathbf{\Omega}_{T_j}^+ \mathbf{\Omega}_{T_j} \mathbf{h}\|_2^q \\ &\geq (1 - \delta_{S+M})^{q/2} \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q - (1 + \delta_M)^{q/2} \sum_{j \geq 2} \|\mathbf{\Omega}_{T_j}^+\|_2^q \|\mathbf{\Omega}_{T_j} \mathbf{h}\|_2^q \\ &\geq (1 - \delta_{S+M})^{q/2} \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q - \sigma_{\min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} M^{q/2-1} (\|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q + 2\eta) \\ &\geq (1 - \delta_{S+M})^{q/2} \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q - \sigma_{\min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} M^{q/2-1} (S^{1-q/2} \|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q + 2\eta) \end{aligned}$$

The second inequality is based on the fact that  $\|\cdot\|_2^q$  satisfies the triangle inequality for any  $0 < q \leq 1$ . While the third inequality uses the D-RIP property.

■

**Lemma 6** Assume that the analysis dictionary  $\mathbf{\Omega}$  has full column rank. If the condition number of  $\mathbf{\Omega}$  satisfies  $\kappa = \sigma_{\max}(\mathbf{\Omega})/\sigma_{\min}(\mathbf{\Omega}) < (S/M)^{1/2-1/q}$ , then the following holds

$$\|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q \leq \frac{1}{\sigma_{\max}^{-q}(\mathbf{\Omega}) - \sigma_{\min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + \frac{2\sigma_{\min}^{-q}(\mathbf{\Omega})M^{q/2-1}}{\sigma_{\max}^{-q}(\mathbf{\Omega}) - \sigma_{\min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} \eta.$$

**Proof** Since  $\mathbf{\Omega}$  has full column rank,  $\mathbf{\Omega}^+ \mathbf{\Omega} = \mathbf{I}$  holds. Then we have

$$\begin{aligned} \|\mathbf{h}\|_2^q &= \|\mathbf{\Omega}^+ \mathbf{\Omega} \mathbf{h}\|_2^q = \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h} + \sum_{j \geq 2} \mathbf{\Omega}_{T_j}^+ \mathbf{\Omega}_{T_j} \mathbf{h}\|_2^q \\ &\leq \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + \sum_{j \geq 2} \|\mathbf{\Omega}_{T_j}^+ \mathbf{\Omega}_{T_j} \mathbf{h}\|_2^q \\ &\leq \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + \sum_{j \geq 2} \|\mathbf{\Omega}_{T_j}^+\|_2^q \|\mathbf{\Omega}_{T_j} \mathbf{h}\|_2^q \\ &\leq \|\mathbf{\Omega}_{T_{01}}^+ \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + \sigma_{\min}^{-q}(\mathbf{\Omega}) M^{q/2-1} (S^{1-q/2} \|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q + 2\eta). \end{aligned}$$

In addition, we have

$$\sigma_{\max}^{-q}(\mathbf{\Omega}) \|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q \leq \sigma_{\min}^q(\mathbf{\Omega}^+) \|\mathbf{\Omega} \mathbf{h}\|_2^q \leq \|\mathbf{\Omega}^+ \mathbf{\Omega} \mathbf{h}\|_2^q = \|\mathbf{h}\|_2^q.$$

Therefore, we obtain

$$[\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}] \|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q \leq \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + 2\sigma_{min}^{-q}(\mathbf{\Omega})M^{q/2-1}\eta.$$

■

The proof of Theorem 1 can now be derived from the previous lemmas, with the procedure as follows:

**Proof** Combinning the lemma 5 and 6, we get

$$\begin{aligned} (1 - \delta_{S+M})^{q/2} \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q &\leq (2\varepsilon)^q + \sigma_{min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} M^{q/2-1} (S^{1-q/2} \|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q + 2\eta) \\ &\leq (2\varepsilon)^q + 2\sigma_{min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} M^{q/2-1} \eta + \frac{\sigma_{min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} (S/M)^{1-q/2}}{\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q \\ &\quad + \frac{2\sigma_{min}^{-2q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} (S/M)^{1-q/2} M^{q/2-1}}{\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} \eta. \end{aligned}$$

Then we obtain

$$\begin{aligned} \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q &\leq \frac{[\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}] 2^q}{(1 - \delta_{S+M})^{q/2} [\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}] - \sigma_{min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} (S/M)^{1-q/2}} \varepsilon^q \\ &\quad + \frac{2\sigma_{max}^{-q}(\mathbf{\Omega})\sigma_{min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} M^{q/2-1}}{(1 - \delta_{S+M})^{q/2} [\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}] - \sigma_{min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} (S/M)^{1-q/2}} \eta. \end{aligned}$$

Therefore, we get

$$\begin{aligned} \|\mathbf{h}\|_2^q &\leq \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + \sigma_{min}^{-q}(\mathbf{\Omega})M^{q/2-1} (S^{1-q/2} \|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q + 2\eta) \\ &\leq \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + 2\sigma_{min}^{-q}(\mathbf{\Omega})M^{q/2-1} \eta + \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2} \|\mathbf{\Omega}_{T_0} \mathbf{h}\|_2^q \\ &\leq \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + 2\sigma_{min}^{-q}(\mathbf{\Omega})M^{q/2-1} \eta + \frac{\sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}}{\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q \\ &\quad + \frac{2\sigma_{min}^{-2q}(\mathbf{\Omega})(S/M)^{1-q/2} M^{q/2-1}}{\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} \eta \\ &\leq \frac{\sigma_{max}^{-q}(\mathbf{\Omega})}{\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} \|\mathbf{\Omega}_{T_{01}} + \mathbf{\Omega}_{T_{01}} \mathbf{h}\|_2^q + \frac{2\sigma_{max}^{-q}(\mathbf{\Omega})\sigma_{min}^{-q}(\mathbf{\Omega})M^{q/2-1}}{\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} \eta \\ &\leq C_1 2^q \varepsilon^q + \frac{2\sigma_{max}^{-q}(\mathbf{\Omega})\sigma_{min}^{-q}(\mathbf{\Omega})M^{q/2-1}}{\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}} [C_1 (1 + \delta_M)^{q/2} + 1] \eta. \end{aligned}$$

In the above proof, we assume that the following inequality holds

$$(1 - \delta_{S+M})^{q/2} [\sigma_{max}^{-q}(\mathbf{\Omega}) - \sigma_{min}^{-q}(\mathbf{\Omega})(S/M)^{1-q/2}] - \sigma_{min}^{-q}(\mathbf{\Omega})(1 + \delta_M)^{q/2} (S/M)^{1-q/2} > 0.$$

It can be simplified as

$$\delta_M + (\kappa^{-q} \rho^{1-q/2} - 1)^{2/q} \delta_{S+M} < (\kappa^{-q} \rho^{1-q/2} - 1)^{2/q} - 1.$$

Again we assume that

$$(\kappa^{-q} \rho^{1-q/2} - 1)^{2/q} - 1 > 0,$$

which leads  $\kappa < \frac{\rho^{1/q-1/2}}{2^{1/q}}$ . ■



## B: Proof of Proposition 2

**Proof** Let  $L := (\rho + 1)S_1$ ,  $\tilde{l} = \rho^{\frac{1}{2-q}}$ ,  $\tilde{S}_q = \frac{L}{\tilde{l}+1}$ , then

$$\begin{aligned}\delta_{(\tilde{l}+1)\tilde{S}_q} &= \delta_{(\rho+1)S_1} \leq \frac{(\kappa^{-1}\rho^{1/2} - 1)^2 - 1}{(\kappa^{-1}\rho^{1/2} - 1)^2 + 1} = \frac{(\kappa^{-1}\tilde{l}^{1-q/2} - 1)^2 - 1}{(\kappa^{-1}\tilde{l}^{1-q/2} - 1)^2 + 1} \\ &\leq \frac{(\kappa^{-q}\tilde{l}^{1-q/2} - 1)^2 - 1}{(\kappa^{-q}\tilde{l}^{1-q/2} - 1)^2 + 1} \leq \frac{(\kappa^{-q}\tilde{l}^{1-q/2} - 1)^{2/q} - 1}{(\kappa^{-q}\tilde{l}^{1-q/2} - 1)^{2/q} + 1}\end{aligned}$$

As  $\delta_m$  is non-decreasing with respect to  $m$  and the map  $\rho \mapsto \frac{(\kappa^{-1}\rho^{1/2}-1)^2-1}{(\kappa^{-1}\rho^{1/2}-1)^2+1}$  is increasing in  $\rho$  for  $\rho \geq 0$ , we only need choose  $l$  and  $S_q$  such that  $l \geq \tilde{l}$  and  $(l+1)S_q = L$  ( $l, S_q \in \mathbb{N}$ ). Recall that  $S_1 \in \mathbb{N}$  and  $(\rho+1)S_1 \in \mathbb{N}$  ( $\rho \geq 2$ ), thus we have  $\sqrt{2} < \tilde{l} < \rho \leq L-1$ , and  $\rho \in \{\frac{n}{L-n} : n = \lceil \frac{2L}{3} \rceil, \dots, L-1\}$ . Thus there exists  $n^*$  such that

$$\frac{n^* - 1}{L - n^* + 1} < \tilde{l} \leq \frac{n^*}{L - n^*},$$

and  $\frac{n^*}{L-n^*} > 2$ . As a result, we have

$$L - n^* \leq \tilde{S}_q < L - n^* + 1.$$

Therefore we can choose  $l = \frac{n^*}{L-n^*}$  and

$$S_q = \lfloor \tilde{S}_q \rfloor = \left\lfloor \frac{\rho+1}{\rho^{\frac{1}{2-q}} + 1} \right\rfloor S_1$$

such that

$$\delta_{(l+1)S_q} < \frac{(\kappa^{-q}l^{1-q/2} - 1)^{2/q} - 1}{(\kappa^{-q}l^{1-q/2} - 1)^{2/q} + 1}.$$

■

## C: Proof of Lemma 1

For the boundedness of  $\|\mathbf{x}^{(k)}\|_2$ , we first show that  $\|\mathbf{\Omega}\mathbf{x}^{(k)}\|_1$  is bounded. We have

$$\begin{aligned}\|\mathbf{\Omega}\mathbf{x}^{(k)}\|_1 &\leq \sum_{i=1}^p [(\omega_i \cdot \mathbf{x}^{(k)})^2 + \varepsilon^{(k)}]^{1/2} \\ &\leq F(\mathbf{x}^{(k)}, \varepsilon^{(k)})/\lambda \leq F(\mathbf{x}^{(0)}, \varepsilon^{(0)})/\lambda.\end{aligned}$$

Thus we get

$$\|\mathbf{\Omega}\mathbf{x}^{(k)}\|_2 \leq \|\mathbf{\Omega}\mathbf{x}^{(k)}\|_1 \leq F(\mathbf{x}^{(0)}, \varepsilon^{(0)})/\lambda.$$

Since it is assumed that  $\mathbf{\Omega}$  has full column rank, we have

$$\begin{aligned}\|\mathbf{x}^{(k)}\|_2 &= \|(\mathbf{\Omega}^T \mathbf{\Omega})^{-1} \mathbf{\Omega}^T \mathbf{\Omega}\mathbf{x}^{(k)}\|_2 \\ &\leq \|(\mathbf{\Omega}^T \mathbf{\Omega})^{-1} \mathbf{\Omega}^T\|_2 \|\mathbf{\Omega}\mathbf{x}^{(k)}\|_2 \\ &\leq \|(\mathbf{\Omega}^T \mathbf{\Omega})^{-1} \mathbf{\Omega}^T\|_2 F(\mathbf{x}^{(0)}, \varepsilon^{(0)})/\lambda.\end{aligned}$$

For  $F(\mathbf{x}^{(k+1)}, \varepsilon^{(k+1)}) \leq F(\mathbf{x}^{(k)}, \varepsilon^{(k)})$ , we consider

$$\begin{aligned}
F(\mathbf{x}^{(k+1)}, \varepsilon^{(k+1)}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}^{(k+1)}\|_2^2 \\
&\quad + \min_{\eta > 0} \left\{ \frac{\lambda q}{\alpha} \sum_{i=1}^p \left[ \eta_i \left( |\boldsymbol{\omega}_i \cdot \mathbf{x}^{(k+1)}|^\alpha + \varepsilon^{(k+1)\alpha} \right) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{\frac{q}{\alpha - q}}} \right] \right\} \\
&\leq \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}^{(k+1)}\|_2^2 \\
&\quad + \frac{\lambda q}{\alpha} \sum_{i=1}^p \left[ \eta_i^{(k+1)} \left( |\boldsymbol{\omega}_i \cdot \mathbf{x}^{(k+1)}|^\alpha + \varepsilon^{(k+1)\alpha} \right) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{(k+1)\frac{q}{\alpha - q}}} \right] \\
&= \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \right. \\
&\quad \left. + \frac{\lambda q}{\alpha} \sum_{i=1}^p \left[ \eta_i^{(k+1)} \left( |\boldsymbol{\omega}_i \cdot \mathbf{x}|^\alpha + \varepsilon^{(k+1)\alpha} \right) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{(k+1)\frac{q}{\alpha - q}}} \right] \right\} \\
&\leq \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}^{(k)}\|_2^2 \\
&\quad + \frac{\lambda q}{\alpha} \sum_{i=1}^p \left[ \eta_i^{(k+1)} \left( |\boldsymbol{\omega}_i \cdot \mathbf{x}^{(k)}|^\alpha + \varepsilon^{(k+1)\alpha} \right) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{(k+1)\frac{q}{\alpha - q}}} \right] \\
&= \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}^{(k)}\|_2^2 \\
&\quad + \min_{\eta > 0} \left\{ \frac{\lambda q}{\alpha} \sum_{i=1}^p \left( \eta_i \left( |\boldsymbol{\omega}_i \cdot \mathbf{x}^{(k)}|^\alpha + \varepsilon^{(k)\alpha} \right) + \frac{\alpha - q}{q} \frac{1}{\eta_i^{\frac{q}{\alpha - q}}} \right) \right\} \\
&= F(\mathbf{x}^{(k)}, \varepsilon^{(k)}).
\end{aligned}$$

## D: Proof of Theorem 3

Since  $\mathbf{A}$  satisfies the  $\Omega$ -RIP property of order  $2l - p$ , we have

$$\begin{aligned}
\|\bar{\mathbf{x}} - \mathbf{x}^*\|_2 &\leq \frac{1}{\sqrt{1 - \delta_{2l-p}}} \|\mathbf{A}(\bar{\mathbf{x}} - \mathbf{x}^*)\|_2 \\
&= \frac{1}{\sqrt{1 - \delta_{2l-p}}} \|(\mathbf{y} - \mathbf{A}\bar{\mathbf{x}}) - (\mathbf{y} - \mathbf{A}\mathbf{x}^*)\|_2 \\
&\leq \frac{1}{\sqrt{1 - \delta_{2l-p}}} (\|\mathbf{y} - \mathbf{A}\bar{\mathbf{x}}\|_2 + \|\mathbf{y} - \mathbf{A}\mathbf{x}^*\|_2) \\
&= \frac{1}{\sqrt{1 - \delta_{2l-p}}} (\|\mathbf{y} - \mathbf{A}\bar{\mathbf{x}}\|_2 + \|\mathbf{e}\|_2).
\end{aligned}$$

In addition, we have

$$\begin{aligned}
\|\mathbf{y} - \mathbf{A}\bar{\mathbf{x}}\|_2 &\leq \sqrt{2F(\bar{\mathbf{x}}, \bar{\varepsilon})} \leq \sqrt{2F(\mathbf{x}^{(0)}, \varepsilon^{(0)})} \\
&= \sqrt{2\lambda \sum_{i=1}^p [|\boldsymbol{\omega}_i \cdot \mathbf{x}^{(0)}|^\alpha + \varepsilon^{(0)\alpha}]^{q/\alpha}}
\end{aligned}$$

Combining the above two inequalities, we get

$$\begin{aligned} \|\bar{\mathbf{x}} - \mathbf{x}^*\|_2 &\leq \frac{1}{\sqrt{1 - \delta_{2l-p}}} \sqrt{2\lambda \sum_{i=1}^p [|\boldsymbol{\omega}_{i, \mathbf{x}^{(0)}}|^\alpha + \varepsilon^{(0)\alpha}]^{q/\alpha}} \\ &\quad + \frac{1}{\sqrt{1 - \delta_{2l-p}}} \epsilon. \end{aligned}$$